

S Z E M L E

Szélsőséges adatok kizárása kísérleti eredmények értékeléséből

A kísérleti (vizsgálati, megfigyelési) adatokban előfordulnak néha a többitől annyira eltérő szélsőséges értékek, melyeket a kísérletező (adatfelvételező) téves megfigyelésnek, hibának tekint. Az ilyen adatot az értékelésből a józan meggondolás alapján kizárja, mert ez a számított átlagot vagy szórást nagyon eltorzíthatja.

A probléma ott merül fel, hogy hol az a határ, melynél valamely gyanús adatot már ki lehet zárni. Szükség van egy matematikai kritériumra, mellyel a szélsőséges adat kizárása felől, a szubjektív megítélést kiküszöbölve, objektív módszerrel dönthetünk.

Minthogy ilyen problémák a mezőgazdasági kísérletekben és laboratóriumi vizsgálatokban gyakoriak és a magyar mezőgazdasági szakirodalom ezt a kérdést tudomásom szerint még nem tárgyalta, először rövid áttekintést nyújtok az érdeklődőknek az ezirányú matematikai próbák közös alap gondolatáról és a szakirodalomban ismertett módszerek felől, majd a legcélszerűbbnek látszó módszert részletesen ismertetem és három példán bemutatom.

A dolgozatot úgy állítottam össze, hogy azok az olvasók, akiket az elméleti rész kevésbé érdekel, a „gyakorlati alkalmazás” fejezet elolvasásával a módszert alkalmazzák tudják.

A részletesen ismertett módszert a nagyüzemi kísérletek eddig kidolgozott módszerlanának szerves kiegészítéseként javaslom.

A módszerek közös alapelve

Adott vizsgálat n adatát egy normális eloszlású alapsokaság mintájának tekintjük. A mintából az átlaggal (\bar{X}) és a szórással (s) az alapsokaság μ középértékét

és σ szórását becsüljük. A minta egy vagy több adata azonban feltűnően eltér a többi adattól. Kérdés, hogy a szélsőséges adatot (adatokat) a vizsgált alapsokasághoz tartozónak, vagy pedig más alapsokaságból származónak tekintsük-e. Utóbbi esetben a kritikus adatot az alapsokaság középértékének vagy szórásának becsléséből kihagyjuk.

Különböző matematikai módszerek ismeretesek, melyekkel fenti kérdés eldönthető. Ezek lényege a következő:

Ha egy normális eloszlású alapsokaságból igen sok n elemű mintát veszünk, a szélsőséges adatok ellenőrzésére megadott képletből számított érték az esetek 1%-ban, 5%-ban, 10%-ban stb. bizonyos kritikus értéket meghalad. Tehát annak valószínűsége, hogy a kritikus értéknél nagyobbat kapjunk 1%, 5%, 10% stb. A kritikus értékeket táblázatok adják meg.

Amikor adott minta egy vagy több szélsőséges adatát ellenőrizzük, a minta adataiból kiszámítjuk a képlet számszerű értékét. Ha ez nagyobb, mint a táblázatosan megadott kritikus érték, akkor $P = 1\%, 5\%, 10\%$ stb. szignifikancia szinten, aszerint, hogy milyen valószínűségi szinten teszteltünk, azt a döntést hozzuk, hogy a mintánkban levő szélsőséges adat nem származik ugyanabból az alapsokaságból, mint a többi adat. A szélsőséges adatot a további számításból ezért kizárjuk.

A szélsőséges érték statisztikai ellenőrzéséhez a minta adatait nagyság szerint rendezzük; az egyes adatokat $X_1, X_2, \dots, X_{n-1}, X_n$ -vel jelöljük. A mindenkor szélsőséges értéket X_1 jelöli. Aszerint tehát, hogy a szélsőséges érték a legnagyobb vagy a legkisebb, X_1 a legnagyobb vagy legkisebb értéket jelöli, X_n pedig a megfelelő utolsó érték, tehát X_1 -től függően a legkisebb vagy legnagyobb adat.

A módszerek

A szélsőséges érték kritériumának problémáját már Chauvant 1850 körül felvetette, alkalmazható módszereket azonban csak a század második negyedében dolgoztak ki. Ezeket Dixon [2] foglalta össze és hasonlította össze, ezért itt csak képletekkel és irodalmi hivatkozással ismertetem. Az irodalomban tájékoztatásul néhány további a tárgyú dolgozatot is felsoroltam [7, 16, 18].

A módszereket A, B, C, D és E-vel jelölt öt csoportba osztottam. Az egyes csoportokon belüli különböző módszerek képleteit $B_1, B_2, B_3; C_1, C_2; D_1, D_2$ jelzi. A csoportonként zárójelben feltüntetett számok irodalmi hivatkozások.

A) χ^2 próba [10]

$$(1) \quad \chi^2 = \frac{\sum (X_i - \bar{X})^2}{\sigma^2}$$

Ha χ^2 nagyobb mint a táblázatos $\chi^2_{p\%}$ érték, akkor a mintából egy szélsőséges adat kizárható. A vizsgálatot addig végezzük, amíg a χ^2 próba szerint a minta elemei homogén sokaságból származónak tekinthetők.

B) Szélsőséges eltérés vizsgálata (B_1 [6, 10, 11] B_2 [8] B_3 [6, 13, 17])

$$(2) \quad B_1 = \frac{X_1 - \bar{X}}{\sigma}$$

$$(3) \quad B_2 = \frac{X_1 - X_2}{\sigma}$$

$$(4) \quad B_3 = \frac{X_1 - \bar{X}}{s}$$

A kiszámított B_1 és B_2 és B_3 értékeket a számláló előjelétől függetlenül mindig pozitívnak tekintjük.

C) Variációs szélesség vizsgálata (C_1 [9, 14] C_2 [11, 12, 15])

$$(5) \quad C_1 = \frac{w}{\sigma}$$

$$(6) \quad C_2 = \frac{w}{s}$$

w = a variációs szélesség, tehát a minta legnagyobb és legkisebb értéke közötti különbség, természetesen a szélsőséges értéket is beleértve.

D) Módosított F próba [6]

1. egy szélsőséges értékre

$$(7) \quad D_1 = \frac{S_1^2}{S^2}$$

$$\text{ahol } S_1^2 = \sum_{i=2}^n (X_i - \bar{X}_1)^2 \text{ és } \bar{X}_1 = \frac{1}{n-1} \sum_{i=2}^n X_i$$

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 \text{ és } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

2. két szélsőséges értékre

$$(8) \quad D_2 = \frac{S_{1,2}^2}{S^2}$$

$$\text{ahol } S_{1,2}^2 = \sum_{i=3}^n (X_i - \bar{X}_{1,2})^2 \text{ és } \bar{X}_{1,2} = \frac{1}{n-2} \sum_{i=3}^n X_i$$

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 \text{ és } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

E) Dixon r kritériuma [2, 3].

Aszerint, hogy több szélsőséges adatot is figyelmen kívül kívánunk hagyni az ellenőrzött (X_1) adat vizsgálatakor, a képletek a következők szerint alakulnak:

1. Minden adat figyelembevétele

$$(9) \quad r_{10} = \frac{X_1 - X_2}{X_1 - X_n}$$

2. Az utolsó adat (X_n) figyelmen kívül hagyása

$$(10) \quad r_{11} = \frac{X_1 - X_2}{X_1 - X_{n-1}}$$

3. Az utolsó és utolsó előtti adat (X_{n-1} és X_n) figyelmen kívül hagyása

$$(11) \quad r_{12} = \frac{X_1 - X_2}{X_1 - X_{n-2}}$$

4. A második adat (X_2) figyelmen kívül hagyása

$$(12) \quad r_{20} = \frac{X_1 - X_3}{X_1 - X_n}$$

5. A második és utolsó adat (X_2 és X_n) figyelmen kívül hagyása

$$(13) \quad r_{21} = \frac{X_1 - X_3}{X_1 - X_{n-1}}$$

6. A második, utolsó és utolsó előtti adat (X_2 , X_{n-1} , és X_n) figyelmen kívül hagyása

$$(14) \quad r_{22} = \frac{X_1 - X_3}{X_1 - X_{n-2}}$$

A módszerek összehasonlítása

Az A , B_1 , B_2 és C_1 módszerek csak akkor jöhetnek számításba, ha az alapsokaság valódi szórása előre ismeretes. Mezőgazdasági jellegű kísérletekben azonban az alapsokaság valódi szórása általában ismeretlen, fenti négy módszer alkalmazási területe ezért szűk. A négy eljárás Dixon [2] szerint körülbelül egyforma értékű. Legegyszerűbb a C_1 . Ha a többszámítás nem jelent hátrányt, és sok az adat, akkor a B_1 eljárást javasolja.

A B_3 eljárás a D_1 -nek felel meg, mivel

$$D_1 = 1 - \frac{B_3}{n-1}$$

A B_3 és C_2 eljárásnál lényeges, hogy a szórás (s) a szélsőséges értéket tartalmazó mintától függetlenül, tehát más mintából határozzuk meg. Ez nehézkessé teszi.

Igen jó a D_1 és D_2 eljárás. A D_1 eljárás azonban csak akkor alkalmazható, ha a szélsőséges érték az eloszlásnak csak az egyik oldalán fordulhat elő. Ha a minta nagy elemszámú, akkor függetlenül attól, hogy a szélsőséges értékek az eloszlásnak csak egyik vagy mindkét oldalán előfordulhatnak, a D_2 eljárás alkalmazása célszerű.

A D eljárás hátránya, hogy alkalmazása nem helyes oly módon, hogy a próbákat folyamatosan végezve mindig egy adattal többet hagyunk ki egészen addig, amíg már nem kapunk szignifikáns eredményt. Több szélsőséges érték esetén ez a módszer ezért hátrányos.

Az E eljárás előnye, hogy a szórás nem kell kiszámítani, ezért egyszerű és gyors, több szélsőséges adat esetén is alkalmazható. Mivel gyakorlatilag az E eljárás, tehát az r kritérium alkalmazása látszik legcélszerűbbnek, alábbiakban részletesebben ismertetem.

Az r kritérium logikai alapja

Tételezzük fel, hogy 7 laboratóriumi szárazanyagvizsgálat közül 6 adat 19,6% és 21,3% között van, a hetedik ellenben 8,2%. Alkalmazzuk az r_{10} kritériumot (9. képlet)

$$\begin{aligned} r_{10} &= \frac{X_1 - X_2}{X_1 - X_n} = \frac{8,2\% - 19,6\%}{8,2\% - 21,3\%} = \\ &= \frac{-11,4\%}{-13,1\%} = 0,870 \end{aligned}$$

a) Nyilvánvaló, hogy minél kisebb a különbség a 6 adat között (vagyis minél kisebb az alapsokaság szórása), és minél nagyobb a szélsőséges adat eltérése a második adattól, annál jobban közeledik a hányados értéke 1-hez.

Tételezzük fel azonban, hogy hat adatkunk 14,3% és 27,4% között változik, és ellenőrizzük ebben az esetben is a 8,2%-os szélsőséges értéket. Ekkor

$$\begin{aligned} r_{10} &= \frac{X_1 - X_2}{X_1 - X_n} = \frac{8,2\% - 14,3\%}{8,2\% - 27,4\%} = \\ &= \frac{-6,1\%}{-19,2\%} = 0,177 \end{aligned}$$

Minél nagyobb a különbség a 6 adat között (vagyis minél nagyobb az alapsokaság szórása), és minél kevésbé tér el a szélsőséges érték a második adattól, annál kisebb a hányados. Ha a „szélsőséges adat” azonos a másodikkal, tehát végeredményben nincs szélsőséges adat, akkor a hányados értéke 0.

Az r érték tehát csak +1 és 0 között változhat. Minél szélsőségebb adatot vizsgálunk, annál nagyobb az r érték.

b) Nagy szórású alapsokaságból is kiválaszthatunk véletlenül néhány igen közelálló értéket, melyek nem reprezentálják az alapsokaság szórását. Ha kevés elemszámú, pl. 4 mintával dolgozunk, könnyen lehetséges, hogy három adat viszonylag jó egyezést mutat, a negyedik azonban nagyon eltér a többitől, jöllehet azonos sokasághoz tartozik. Minél nagyobb elemszámú mintával dolgozunk, annál kevésbé fordulhat elő ilyen véletlen. Ez az oka annak, hogy minél kisebb elemszámú a minta, annál nagyobb az azonos valószínűségi szinthez tartozó kritikus r érték.

c) Ha növeljük a minta elemszámainak számát, fokozódik annak valószínűsége, hogy több szélsőséges, nem az alapsokasághoz tartozó adat is van a mintában.

Tételezzük fel, hogy 18 mérés közül 15 mérés értéke 18,3% és 24,6% között van, három mérés értéke pedig 8,2%, 11,7% és 30,5%. Ha a 8,2%-os értéket az előbbi r_{10} kritériummal vizsgálánk, akkor

$$\begin{aligned} r_{10} &= \frac{X_1 - X_2}{X_1 - X_n} = \frac{8,2\% - 11,7\%}{8,2\% - 30,5\%} = \\ &= \frac{-3,5\%}{-22,3\%} = 0,157 \quad (P = 5\% \text{ szinten} \\ &\quad r_{10} = 0,313) \end{aligned}$$

E próbából levont következtetésünk, hogy a 8,2%-os érték az összes többi adattal

azonos alapsokasághoz tartozik, aligha lenne helytálló. De ugyanezt a próbát elvégezve a 30,5%-os értékre

$$r_{10} = \frac{X_1 - X_2}{X_1 - X_n} = \frac{30,5\% - 24,6\%}{30,5\% - 8,2\%} = \frac{5,9\%}{22,3\%} = 0,265 \quad (P = 5\% \text{ szinten } r_{10} = 0,313)$$

hányadosnak sincs bizonyító ereje a 30,5 százalékos érték kizárására.

Ilyen esetekben indokolt az ellenőrzött értéken kívüli gyanús adatok figyelmen kívül hagyása. Alkalmazzuk a 8,2 %-os adat ellenőrzésére az r_{21} kritériumot, a 30,5 %-os adat ellenőrzésére a r_{12} kritériumot.

$$r_{21} = \frac{X_1 - X_3}{X_1 - X_{n-1}} = \frac{8,2\% - 18,3\%}{8,2\% - 24,6\%} = \frac{-10,1\%}{-16,4\%} = 0,616 \quad (P = 5\% \text{ szinten } r_{21} = 0,440)$$

$$r_{12} = \frac{X_1 - X_2}{X_1 - X_{n-2}} = \frac{30,5\% - 24,6\%}{30,5\% - 18,3\%} = \frac{5,9\%}{12,2\%} = 0,484 \quad (P = 5\% \text{ szinten } r_{12} = 0,377)$$

Ezzel az eljárással tehát mindkét érték kizárásának indokoltsága igazolható.

A 8,2%-os érték kizárása után r_{11} kritériummal ellenőrizhető a 11,7%-os érték kizárásának jogosultsága is, és így folytatólagosan ellenőrizhetők a szélsőséges értékek egészen addig, amíg már nem kapunk szignifikáns eredményt.

d) Végül még meggondolandó a következő. A célszerű r kritérium megválasztása a c) pontban foglaltak szerint kétségtelenül kissé önkényes, szubjektív megítélésre ad lehetőséget, nem tekinthető tehát teljesen objektív módszernek. Ezért indokolt, a minta elemszámától függően, a különböző r kritériumok adott mintától független, tehát előzetesen meghatározott kombinált alkalmazása. Az r kritériumok kombinált alkalmazására összeállított táblázatot Dixon-tól [4, 5] átvetem, és a 6. táblázatban ismertetem.

Gyakorlati alkalmazás

1. A talaj N tartalmának változását vizsgáljuk 2 hetes időközökben vett mintákkal. A kijelölt területen minden vizs-

gálat alkalmával 6 talajmintát veszünk. Ezekben egyenként meghatározzuk az oldható N -t, mg/100 g talajban kifejezve, és a kapott adatokat átlagoljuk.

A laboratóriumban meghatározott N értékeket az 1. táblázat mutatja. Az egy alkalommal vett minták N értékének átlagát az „Átlag” sor mutatja.

1. táblázat

Két hetenként 6—6 talajmintából meghatározott oldható N mg/100 g talaj adatok

	Mintavétel időpontja		
	I	II	III
	2,5	2,4	3,0
	1,7	2,0	2,7
	1,8	0,7	2,6
	1,4	3,0	3,6
	4,6	2,9	3,0
	2,1	2,4	3,5
Összesen	14,1	13,4	18,4
Átlag	2,35	2,23	3,07
Korrigált összeg ..	9,5	12,7	18,4
Korrigált átlag ...	1,90	2,54	3,07

Az adatok varianciaanalízise a 2. táblázatban nem igazol különbségeket a 2 hetenként vett minták átlaga között. Tehát nem bizonyíthatjuk, hogy a talaj N tartalma változott a három vizsgálati időpontban. (A kísérleti hiba variációs koefficiense, $CV = 34\%$.)

2. táblázat

Az eredeti 18 adatból számított varianciaanalízis

Tényező	SQ	FG	MQ
Összes	13,55	17	
Mintavétel időpontja	2,45	2	1,225
Hiba	11,10	15	0,740

Az I. és II. vizsgálati időpontban azonban a 4,6 és 0,7-os adat „gyanús”. Vizsgáljuk meg az r kritériummal, hogy ezeket „hibás” adatként ki kellene-e hagyni. Példánkban az átlagokat 6 adatból határoztuk meg, ezért a 6. táblázatból

$$n = 6 \text{ sorban } r_{10} = \frac{X_1 - X_2}{X_1 - X_n}$$

képlettel vizsgáljuk meg a két adat kizárásának jogosultságát.

Minthogy X_1 jelenti a vizsgált szélsőséges adatot, X_2 nagyság szerint a szomszéd adatot, végül X_n az X_1 -től nagyság szerint legtávolabbi adatot, az I. mintavételi időpontban

$$r_{10} = \frac{4,6 - 2,5}{4,6 - 1,4} = \frac{2,1}{3,2} = 0,656$$

nagyobb, mint az $n = 6$ sorban a $P = 2\%$ szintre megadott 0,644-es érték. A 4,6-os adatot $P = 2\%$ szinten kizárjuk.

A II. mintavételi időpontban a 0,7 adat vizsgálata:

$$r_{10} = \frac{0,7 - 2,0}{0,7 - 3,0} = \frac{-1,3}{-2,3} = 0,565$$

nagyobb, mint az $n = 6$ sorban a $P = 5\%$ szintre megadott 0,560-es érték. A 0,7-es adatot $P = 5\%$ szinten kizárjuk.

A két kizárt adat nélküli összegeket és átlagokat az 1. táblázat „korrigált összeg” és „korrigált átlag” sora mutatja. Utóbbi sorból kitűnik, hogy a kéthetenkénti átlagadatok az eredeti átlagokhoz viszonyítva lényegesen megváltoztak.

A két szélsőséges adat nélküli varianciaanalízis (3. táblázat) $P = 1\%$ szinten igazolja, hogy a vizsgált időpontokban a talaj N tartalma változott, előbbi következtetésünk tehát téves lett volna.

A szélsőséges adatok kizárásának eredményét mutatja az is, hogy előbbi 34%-os CV érték a két adat kihagyása után 16%-ra csökkent.

3. táblázat

A két szélsőséges adat nélkül számított varianciaanalízis

Tényező	SQ	FG	MQ
Összes	5,92	15	
Mintavétel időpontja	3,71	2	1,760**
Hiba	2,21	13	0,170

** $P = 1\%$

2. Laboratóriumban paralel vizsgálattal egy új fehérjemeghatározási módszer hibáját határozzuk meg 8 különböző mintából.

A 8 minta paralelvizsgálatának eredményét a 4. táblázat tartalmazza.

4. táblázat

Nyolc paralel fehérjevizsgálat eredménye

Minta sorszáma	Fehérje %		Különbség paralelek között
	a	b	
	paralelek		
1	18,3	18,4	—0,1
2	19,6	20,1	—0,5
3	23,1	22,1	+1,0
4	20,3	20,6	—0,3
5	14,2	14,0	+0,2
6	18,6	18,4	+0,2
7	19,5	19,9	—0,4
8	26,2	25,9	+0,3

A kérdés az, hogy e paralelvizsgálatokban mutatkozó különbség alapján ezzel a módszerrel mekkora vizsgálati hibára kell számítanunk. Ha mind a 8 adatot figyelembe vesszük, akkor a vizsgálati módszer hibaszórása

$$s = \sqrt{\frac{(-0,1)^2 + (-0,5)^2 + \dots + 0,3^2}{8}} = \sqrt{0,21} = 0,46$$

Kérdéses azonban, hogy a 3. számú mintában mutatkozó +1,0-os különbség a vizsgálati módszer normál hibájának tekinthető-e, vagy pedig egyéb véletlen hibának, tehát az új módszer vizsgálati hibaszórásának meghatározásából ki kell hagyni.

A statisztikai vizsgálat szempontjából nem a laboratóriumi számozott mintákat, hanem a 8 paralel vizsgálat közötti különbséget tekintjük az új vizsgálati módszerrel előforduló hibák alapsokaságából vett 8 elemű mintának, melyből a vizsgálati hibák valódi szórását kívánjuk becsülni.

A 6. táblázatban az $n = 8$ sornak megfelelően az r_{11} képlettel ellenőrizzük a +1,0-os értéket:

$$r_{11} = \frac{X_1 - X_2}{X_1 - X_{n-1}} = \frac{1,0 - 0,3}{1,0 + 0,4} = \frac{0,7}{1,4} = 0,500$$

nagyobb, mint 0,479, de kisebb, mint 0,554, tehát $P = 5\%$ szinten még nem, de $P = 10\%$ szinten állíthatjuk, hogy a +1,0 érték véletlen hiba, és nem tekinthető az új módszerrel együttjáró hibának.

Általában ha az r kritérium $P = 5\%$ szinten szignifikanciát igazol, a kritikus adatot hagyjuk ki. Ha $P = 10\%$ szinten igazol csak szignifikanciát, és a vizsgálat nem ismételhető meg, a kritikus adatot ebben az esetben is hagyjuk ki. Ha a vizsgálat megismételhető, akkor $P = 10\%$ -os szignifikancia esetén nagyobb elemszámú mintából ismét végezzük el a vizsgálatot.

Példánkban a kritikus $+1,0$ érték kihasználása után az új módszer hibaszórása $0,31$, tehát némileg kisebb, mintha mind a 8 adattal számolunk.

3. A szélsőséges értékek különös problémát jelentenek nagyüzemi kísérletekben. Itt gyakori eset, hogy néhány gazdaságban a többi gazdaság adataitól teljesen eltérő kezeléskülönbségek vannak. Ennek nyilvánvaló kísérleti hiba is lehet oka, amit a gazdaság vagy a kísérleteket irányító központ megbízottja a vegetációs idő alatt észrevesz. Ebben az esetben a kérdéses adatok minden matematikai ellenőrzés nélkül kizárhatóak. Előfordulhatnak azonban olyan esetek, amikor semmilyen szakmai támpont nincsen, ami a szélsőséges értékek kizárását indokoltá tenné. A nagyüzemi kísérletek jellege miatt egy vagy két szélsőséges ellentmondás miatt nem zárhatjuk ki az összes gazdaság adatát, mert egy évet vesztenénk, és adott év termelési körülményei nem ismételhetők meg.

A kezeléskülönbséget és a szignifikáns különbséget azonban egy vagy két szélsőséges adat nagyon eltorzíthatja, és ezért ezek kizárásának elbírálására szükséges egy minden szubjektivitástól mentes, objektív kritérium.

A szélsőséges, ellentmondó adat nemcsak hibából származhat, hanem utalhat arra is, hogy a kezeléskülönbség adott gazdaság termelési körülményei között, ismeretlen ok miatt változott meg a többi gazdaság eredményéhez viszonyítva. Mint-hogy azonban a szignifikáns különbség kiszámításakor előfeltétel a különbségek legalábbis megközelítő normális eloszlása, az ilyen adatot mindenképpen ki kell zárni az értékelésből.

Nagyüzemi kísérletek értékelésekor ezért célszerűnek látszik, hogy minden olyan gazdaságot, melyben a kezeléskülönbség a többi gazdaságban mutatkozó kezeléskülönbségtől $P = 10\%$ szinten szélsőségesen eltérőnek mutatkozik, az értékelésből kizárjuk.

Példa: búzával végzett nagyüzemi kísérletben 14 gazdaságból két kezelésre vonatkozólag az 5. táblázatban ismertetett adatokat kapjuk.

5. táblázat

Nagyüzemi kísérlet búzatermés adatai
q/kh-ban, 14 gazdaságból, két kezeléssel

Gazdaság	Kezelések		Terméskülönbség q/kh (diff)	
	A	B		
	termés q/kh		+	-
1	16,4	13,8	2,6	
2	11,3	9,6	1,7	
3	18,8	12,7	6,1	
4	17,6	15,1	2,5	
5	12,7	13,4		0,7
6	13,8	11,5	2,3	
7	16,9	17,5		0,6
8	19,7	15,3	4,4	
9	12,8	18,1		5,3
10	13,2	12,1	1,1	
11	13,3	11,7	1,6	
12	14,4	15,4		1,0
13	7,7	7,5	0,2	
14	13,9	13,4	0,5	
Összesen	202,5	187,1	23,0	7,6
Átlag	14,5	13,4	diff = 15,4	

Mind a 14 adat figyelembevételével a két kezelés közötti átlagos terméskülönbség $\frac{15,4 \text{ q/kh}}{14} = 1,1 \text{ q/kh}$.

A terméskülönbségek szórásnégyzetét

$$(15) \quad s_{\text{diff}}^2 = \frac{\sum(\text{diff})^2 - \frac{(\sum \text{diff})^2}{n}}{n - 1}$$

és a $P = 5\%$ -ra számított szignifikáns differenciát

$$(16) \quad SzD_{5\%} = t_{5\%} \sqrt{\frac{s_{\text{diff}}^2}{n}}$$

képletekkel meghatározva, példánkban

$$s_{\text{diff}}^2 = \frac{111,76 - \frac{(15,4)^2}{14}}{13} = 7,29$$

és

$$SzD_{5\%} = 2,16 \sqrt{\frac{7,29}{14}} = 1,56 \text{ q/kh}$$

Mínt hogy a $SzD_s\% = 1,56$ q/kh, a két kezelés közötti 1,1 q/kh-s terméskülönbség nem tekinthető szignifikánsnak.

Vizsgáljuk meg a különbség szórásának és $SzD_s\%$ értékének százalékos értékét is a két kezelés átlagára (13,8 q/kh) vonatkoztatva.

$$(17) \quad s_{diff}\% = \frac{\sqrt{s_{diff}^2} \cdot 100}{\text{két kezelés termésátlag}}$$

$$\text{képlettel, példánkban } s\% = \frac{\sqrt{7,29} \cdot 100}{13,8} = 19,6\%$$

$$\text{és a } SzD_{s\%-os} \text{ értéke} = \frac{1,56}{13,8} = 11,3\%.$$

Az igen nagy $s_{diff}\%$, de a százalékban kifejezett $SzD_s\%$ is arra utal, hogy a két kezelés közötti különbségnek igen nagy a hibája.

Az 5. táblázat adatait vizsgálva, a 3., 8. és 9. gazdaságokban mutatkozó terméskülönbségek nagyon eltérnek a többtől. Ellenőrizzük ezeket az r kritériummal. A 3. gazdaságban a 6,1 q/kh-s terméskülönbség ellenőrzésére alkalmazzuk 6. táblázat $n = 14$ sorához előírt r_{22} kritériumot.

$$r_{22} = \frac{X_1 - X_3}{X_1 - X_{n-2}} = \frac{6,1 - 2,6}{6,1 + 0,6} = \frac{3,5}{6,7} = 0,522$$

$P = 10\%$ szinten szignifikanciát igazol. A 3. gazdaság adatait ezért zárjuk ki a további vizsgálatból.

A 8. gazdaságban +4,4 q/kh a terméskülönbség. Mínt hogy a 3. gazdaság kizárása után csak 13 gazdaságunk maradt, alkalmazzuk a r_{21} kritériumot.

$$r_{21} = \frac{X_1 - X_3}{X_1 - X_{n-1}} = \frac{4,4 - 2,5}{4,4 + 1,0} = \frac{1,9}{5,4} = 0,352$$

$P = 10\%$ szinten nem szignifikáns. A 8. gazdaságot tehát nem zárjuk ki az értékelésből.

A 9. gazdaságban -5,3 q/kh a terméskülönbség. Mivel az előzőekben csak a 3. gazdaságot zártuk ki, ismét az $n = 13$ -ra megadott r_{21} kritériumot alkalmazzuk.

$$r_{21} = \frac{X_1 - X_3}{X_1 - X_{n-1}} = \frac{-5,3 + 0,6}{-5,3 - 2,6} = \frac{-4,7}{-7,9} = 0,595$$

még $P = 2\%$ szintet is meghaladón szignifikáns. A 9. gazdaságot tehát szintén kizárjuk a további értékelésből.

A megmaradt 12 gazdaság átlagában az A kezelés átlagtermése 14,2 q/kh, a B kezelés átlagtermése 13,0 q/kh, a két kezelés közötti átlagos különbség 1,2 q/kh. Lényegesen tehát nem változott a 14 gazdaságból számított 1,1 q/kh terméskülönbséghez viszonyítva. A szórásnégyzet azonban

$$s_{diff}^2 = \frac{46,46 - \frac{(14,6)^2}{12}}{11} = 2,609$$

és a $P = 5\%$ -ra számított szignifikáns különbség

$$SzD_s\% = 2,20 \sqrt{\frac{2,609}{12}} = 1,0$$

Mínt hogy az újonnan kiszámított $SzD_s\%$ kisebb, mint az átlagos kezeléskülönbség, a két kezelés közötti különbség $P = 5\%$ szinten szignifikáns. Az r kritérium alkalmazása után tehát bizonyítottnak tekinthető, hogy a két kezelés között különbség van.

Százalékban kifejezve a szórás és a szignifikáns különbséget (a két kezelés átlagtermése 12 gazdaságban 13,6 q/kh)

$$s_{diff}\% = \frac{\sqrt{2,609} \cdot 100}{13,6} = 11,8\%$$

és

$$SzD_{s\%}\% = \frac{1,0}{13,6} = 7,3\%$$

Ezek az értékek a 14 adatból számítottakkal összehasonlítva szintén azt mutatják, hogy az r kritérium alkalmazásával a kísérlet eredményét lényegesen pontosabbá tettük.

E példából esetleg arra lehetne gondolni, hogy az r kritérium csak a szórás vagy a szignifikáns különbség csökkentésére alkalmas. Az α példából azonban már kitűnt, hogy a középértékeket is pontosabban becsüljük, ha a szélsőséges érték a mintának csak „egyik oldalán” van. Így nagyüzemi példánkban is, ha történetesen eredetileg csak 13 adatunk lett volna, a 3. gazdaság nélkül, akkor a két kezelés közötti terméskülönbség nem szignifikánsan 0,7 q/kh lenne, holott a helyes becslés, mint láttuk 1,2 q/kh, és ez a különbség szignifikáns.

Az r kritériumra alkalmazott képletek kitöltésével kapcsolatban még megjegyzendő, hogy az előjeleket igen óvatosan kell kezelni. Ha ugyanis a képletben a negatív (második) tag helyére negatív

6. táblázat

Kritikus értékek szélsőséges adatok kizárására Dixon r kritériumával (Átvéve Dixon és Massey F. [5] könyvéből)

Képletek	Adatok száma n	Valószínűségi szint (P)				
		10%	5%	2%	1%	0,5%
$r_{10} = \frac{X_1 - X_2}{X_1 - X_n}$	3	0,886	941	976	988	994
	4	679	765	846	889	926
	5	557	642	729	780	821
	6	482	560	644	698	740
	7	434	507	586	637	680
$r_{11} = \frac{X_1 - X_2}{X_1 - X_{n-1}}$	8	479	554	631	683	725
	9	441	512	587	635	677
	10	409	477	551	597	639
$r_{21} = \frac{X_1 - X_3}{X_1 - X_{n-1}}$	11	517	576	638	679	713
	12	490	546	605	642	675
	13	467	521	578	615	649
$r_{22} = \frac{X_1 - X_3}{X_1 - X_{n-2}}$	14	492	546	602	641	674
	15	472	525	579	616	647
	16	454	507	559	595	624
	17	438	490	542	577	605
	18	424	475	527	561	589
	19	412	462	514	547	575
	20	401	450	502	535	562
	21	391	440	491	524	551
	22	382	430	481	514	541
	23	374	421	472	505	532
	24	367	413	464	497	524
	25	360	406	457	489	516

adatot helyettesítünk, akkor ez pozitív előjelű lesz. Az r kritérium kiszámításakor ezért pontosan tegyük ki az előjeleket.

Végül meg szeretném még jegyezni, hogy az ismertett módszerek közül jelenlegi formájában egyik sem alkalmas 2-nél több kezelés szántóföldi kísérletekben valamely parcella adatának kizárására. Ebben az esetben a kizárás a kísérletező szubjektív ítéletétől függ, a kizárt parcella adatát a hiányzó parcellák módszerével helyettesítjük [1, 19].

Összefoglalás

Mezőgazdasági kísérletekben, adatfeldvételezésekben, vagy laboratóriumi vizsgálatokban gyakran előfordulnak szélsőséges értékek, melyek az átlag vagy szórás (szignifikáns különbség) kiszámításakor az

értékelés eredményét torzítják. Dixon [2] után röviden ismertetjük a szélsőséges adatok objektív matematikai módszerrel történő kizárására szolgáló különböző eljárásokat. Részletesen Dixon [2, 3, 4] által javasolt r kritérium alkalmazását ismertetjük, és azt 3 példán bemutatjuk.

Különös részletességgel tárgyaljuk a nagyüzemi kísérletekben egyes gazdaságok által közölt, a többinek ellentmondó adatok kizárását. *Javasoljuk, hogy a nagyüzemi kísérletekben a szélsőséges adatok kizárására az r kritériumot $P = 10\%$ -os szinten alkalmazzuk, és a módszert a nagyüzemi kísérletek eddig kidolgozott módszertanának szerves kiegészítésének tekintsük.*

SVÁB JÁNOS

Érkezett: 1959. szeptember 21.

Irodalom

- [1] *Cochran, W. G.* : Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, **3**, 22—38. 1947.
- [2] *Dixon, W. J.* : Analysis of extreme values. *Annals Math. Stat.* **21**, 488—506. 1950.
- [3] *Dixon, W. J.* : Ratios involving extreme values. *Annals Math. Stat.* **22**, 68—78. 1951.
- [4] *Dixon, W. J.* : Processing data for outliers. *Biometrics*, **9**, 74—89. 1953.
- [5] *Dixon, W. J. & Massey, F. J.* : Introduction to statistical analysis Mc Graw-Hill. New York. 1957.
- [6] *Grubbs, F. E.* : Sample criterion for testing outlying observations. *Annals Math. Stat.* **21**, 27—58. 1950.
- [7] *Hartley, H. O.* : The range of random samples. *Biometrika*, **32**, 301—310. 1942.
- [8] *Irwin, J. O.* : On the criterion for a rejection of outlying observations. *Biometrika*, **17**, 238—250. 1925.
- [9] *Joyce, M. May* : Extended and corrected tables of the upper percentage points of the „Studentized” Range”. *Biometrika*, **39**, 189—191. 1952.
- [10] *McKay, A. T.* : The distribution of the difference between the extreme observation and the sample mean of n from a normal universe. *Biometrika*, **27**, 466—471. 1935.
- [11] *Nair, K. R.* : The distribution of the extreme deviate from the sample mean and its studentized forms. *Biometrika*, **35**, 118—144. 1948.
- [12] *Newman, D.* : The distribution of ranges in samples from a normal population, expressed in terms of an independent estimate of the standard deviation. *Biometrika*, **31**, 20—30. 1940.
- [13] *Pearson, E. S. & Chandra Sekar* : The efficiency of statistical tools and a criterion for a rejection of outlying observations. *Biometrika*, **28**, 308—320. 1936.
- [14] *Pearson, E. S. & Hartley, H. O.* : The probability integral of the range in samples of n observations from the normal population. *Biometrika*, **32**, 301—310. 1942.
- [15] *Pearson, E. S. & Hartley, H. O.* : Tables of the probability integral of the studentised range. *Biometrika*, **33**, 89—99. 1942.
- [16] *Rider, P. R.* : Criteria for rejection of observations. Washington University Studies — New Series: Science and Technology, No. 8. St. Louis. 1933.
- [17] *Thompson, W. R.* : On the criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Annals Math. Stat.* **6**, 214—219. 1935.
- [18] *Tippet, L. H. C.* : The extreme individual and the range of samples taken from a normal population. *Biometrika*, **17**, 151—164. 1925.
- [19] *Yates, E.* : The analysis of replicated experiments when the field results are incomplete. *Empire J. Exp. Agric.* **1**, 129—142. 1933.